Dutch Cooking with xAPI Recipes

The Good, the Bad, and the Consistent

Alan Berg
ICT-Services
University of Amsterdam
Amsterdam, The Netherlands
a.m.berg@uva.nl

Maren Scheffel, Hendrik Drachsler, Stefaan Ternier, Marcus Specht
Welten Institute, Open Universiteit
Heerlen, The Netherlands
{msf, hdr, str, spe}@ou.nl

***Abstract*—This short paper presents the experiences of several Dutch projects in their application of the xAPI standard and different design patterns including the deployment of Learning Record Stores. In this paper we share insights and argue for the formation of an international Special Interest Group on interoperability issues to contribute to the Open Analytics Framework as envisioned by SoLAR and enacted by the Apereo Learning Analytics Initiative. Therefore, we provide an overview of the advantages and disadvantages of implementing the current xAPI standard by presenting projects that applied xAPI in very different ways followed by the lessons learned.**

Keywords- learning analytics; xAPI; data standardization; learning record store; data silos

I. INTRODUCTION

The Experience API (xAPI) – formerly known as TinCan API – was publicly launched in April 2012. Since 2014 numerous projects and initiatives in Europe have been applying the xAPI specification as a metadata approach to securely aggregate learning events ready for digestion by Learning Record Stores (LRSs) and analytics engines. There are three innovative aspects about xAPI that are appealing to digital education providers: it is (1) learner activity centered, (2) system independent, (3) straightforward to implement.

However, the xAPI universe is not perfect. There are also issues that need to be addressed by its adopters. For instance, the application of xAPI does not solve all interoperability problems. In a white paper about inter-LRS communication, Downes, Shahrazad and Smith [1] suggest that the transfer of xAPI data from one LRS to another does not always work smoothly. This highlights the importance of the consistent application of the xAPI recipes and the validation of statements according to the recipe. Another related issue is the lack of shared conventions and best practice examples for xAPI statements that are authoritatively endorsed by educational communities. This raises a level of uncertainty for adopters over their own xAPI definition and approaches.

This paper addresses these shortcomings. The different projects mentioned in this paper have been collaborating through the Special Interest Group (SIG) on Learning Analytics[1] within the Dutch umbrella organization for higher education SURF. We propose an emerging xAPI standard and provide best practice examples and lessons learned for other adopters of the xAPI specification. We wish to extend our approach with other xAPI recipes cook books [2] and collaboration with those working on similar standards such as Caliper [3]. Our aim is to nucleate a wider community conversation on applying recipes to insure maximum data coherency within the wider educational ecosystems and specifically for the learning analytics community.

II. THE PROJECTS

In June 2012 the University of Amsterdam (UvA) initiated a stimulus project for learning analytics known as UvAInform. The focus of the project was to expand the local understanding and evidence for how learning analytics could impact the universities' educational processes. The project included seven pilots mostly centered on dashboard building and a generic infrastructure component, a UvA-developed LRS named Larissa[2]. From 2012 to 2014, the central services of UvA invested in instrumenting open source xAPI connectors for the Sakai LMS and the Apereo Open Academic Environment (OAE)[3]. The aim was to generate wider adoption by researchers. Researcher involvement was seen as a key factor in understanding and developing learning analytic services.

In early 2014, the Open University of the Netherlands received funding for the European project ECO[4] to develop a single entry portal for various MOOC providers and to create a learning analytics infrastructure [4]. The ECO project is comprised of a set of learning platforms that have their own logging and monitoring systems. Each platform can use its proprietary methodology as long as it also provides the required data according to the xAPI specification. The established LRS architecture with xAPI statements then allows for the homogenous calculation of learning analytics indicators for the partner platform.

Another European project, the Learning Analytics Community Exchange (LACE)[5], collects and visualizes evidences to support learning analytics best practices for the school, workplace and higher education sector. Within the LACE project the Open University of the Netherlands conducts experimental studies focused on educational evaluation of advanced analytics tools. Among mobile

[1] https://www.surfspace.nl/sig/18-learning-analytics/

[2] https://github.com/Apereo-Learning-Analytics-Initiative/Larissa
[3] http://oaeproject.org
[4] https://ecolearning.eu
[5] http://www.laceproject.eu

learning analytics, BioFeedback and environmental data are used to identify conditions for productive and unproductive learning contexts. In the Learning Pulse study [5], data from different sources, i.e. interaction data from the PCs, heart rate and step count from a fitness tracker, weather data and contextual data such as noise level, are stored as xAPI statements in an LRS.

III. xAPI

The xAPI specification[6] is used to collect learners' digital traces. It registers who performs which activity with which object at which time and in which context. xAPI statements take the form actor-verb-object to store an experience. In a context object additional details can be described, e.g. the geo location, the course an experience is related to, etc. xAPI statements can also capture the outcome of an experience, e.g. the result of a quiz or the answer to a question, etc. A detailed example can be found on the xAPI website[7].

Unlike SCORM (Sharable Content Object Reference Model) [6], where the users and their activities are one entity hidden among others, and CAM (Contextualized Attention Metadata) [7], which is event-centered, xAPI positions a user's actions at the center of attention. This enables didactical designs and learning processes that can be directed towards personalized learning [8]. This approach has also been followed by IMS Caliper [3]. Several standards have been developed for storing learner events in a specific data format [9], however, xAPI has taken a leading role mainly due to its ability to provide a technical solution to store the activity events in Learning Record Stores (LRSs) [10]. The xAPI specification's success also stems from the low threshold of effort to getting started with collecting xAPI events which is due to xAPI's simplicity: institutions can define their own xAPI statements. The ADL consortium has provided definitions[8,9] as guidelines to design standard-conformant xAPI statements. In addition, an adopter can also use the standards and contribute their own definitions for events outside the scope of the core set. The freedom of definition and the focus on learning experiences as the center of the learning metadata definition makes xAPI malleable in a complex online education-centered ecosphere. The xAPI approach offers the advantage of system independence. This enables practitioners to aggregate traces from learners over various devices and systems by collecting those traces centrally. The cleaning and mangling of data has been a significant if not the most significant cost to analytics projects [11]. With xAPI, however, the collected data are stored securely and centrally in an LRS in a known and machine-readable format and are searchable through standard web services. Applying xAPI methodology across an organization enables the creation of a consistent analytics infrastructure. Therefore, increasing adoption[10] of the xAPI standard across the educational sector also increases the opportunity for sectorwide plug and play analytics services.

One of the advantages of xAPI, i.e. its openness when it comes to designing the xAPI statements, is also one of its disadvantages [12]. This holds especially true when trying to combine data collected in different platforms. The interoperability issue is not a new one and has been described for other standards such as IMS LD [13] and SCORM [6] long before xAPI was created. The recently published Edinburgh Statement for Learning Analytics Interoperability[11] advocates open dialogue between vendors, practitioners and organizations and the establishment of common, shared and accessible spaces to facilitate this dialogue.

Several sources of xAPI recipes are currently available: the ADLnet website[12] offers a library of statements and Kitto et al. [2] have published their collection on Github[13]. For the three projects mentioned above an overview of all xAPI statements implemented to enable the deployment of customized learning analytics dashboards has been developed and is available online[14]. This inter-project and inter-institutional specification of xAPI aims to stimulate a joint collection of xAPI data within the Netherlands and to also contribute to the definition of xAPI specifications and their usage around the world [14].

IV. Outcomes and Lessons Learned

Both UvAInform and ECO showed that a centralized approach to gathering and querying student activities is needed. There are so many heterogeneous IT systems that getting a complete picture of the students' behavior is hardly possible. This experience strengthens the argument for deployment of centralized data collection. Both ECO and UvAInform are built on a more traditional LMS architecture, where xAPI services are hosted for a given identity provider (IDP). Well known examples of identity providers are websites that allow users to log in with their Google, Twitter, Facebook, or OpenID credentials so that those services act as an identity provider. Within ECO an IDP component was created that manages all ECO user accounts. The user object for every ECO xAPI statement therefore relates to a valid ECO account. Such an architectural configuration calls for an xAPI binding that binds the user identifier in the ECO IDP to the actor object. An advantage of this binding is that an ECO user can alter all information (including his/her email address) without breaking the link to already stored xAPI statements. Another advantage is, that an extra level of anonymity is added to the user data.

In ECO the MOOC providers submit xAPI statements to a cloud-based xAPI proxy. The proxy component takes care of managing a copy of the xAPI statements. Furthermore, it forwards the statements to a BigData LRS, a customized data store optimized for making scalable and performant queries. These components were introduced to enable queries by the

[6] https://github.com/adlnet/xAPI-Spec
[7] https://experienceapi.com/statements-101/
[8] http://xapi.vocab.pub/datasets/adl/activities/ and http://xapi.vocab.pub/datasets/adl/verbs/
[9] https://registry.tincanapi.com/#home/verbs and https://registry.tincanapi.com/#home/activityTypes
[10] http://experienceapi.com/adopters/
[11] http://bit.ly/Edi_statement
[12] http://experienceapi.com/recipes/
[13] https://github.com/kirstykitto/CLRecipe
[14] http://bit.ly/DutchXAPIreg

dashboard component that cannot be executed live on the LRS due to long reaction times. Introducing the xAPI proxy is beneficial for error handling and caching. In case the LRS is down, or cannot handle the load, the proxy can publish statements at a later moment. The ECO LRS does not have direct query access to the BigData LRS which is why the proxy was added between the dashboard and the BigData LRS. The proxy is responsible for executing queries and generating cached reports that are then forwarded to the dashboard to be visualized to the users. As a result, the dashboard never has access to the raw data.

Within the UvAInform project an Extract Transform Load (ETL) layer was needed to harvest the data from the different source systems. Within the ETL layer, xAPI recipe consistency is enforced through a common set of transformations. To strengthen this instrument and to maintain a consistent usage an authority needs to own and curate the transformations. One potential authority is the Apereo Learning Analytics Initiative that can publish the transformations to its community context in combination with a SoLAR[15] Work Group to curate.

An xAPI infrastructure also needs to take into account expected future demands. One of these demands will be the collection of data from fitness trackers. As was done in LACE's Learning Pulse study, it took little effort to place these data in the LRS via a number of xAPI recipes. These xAPI recipes should, however, take into account the uncertainty in sensor measurements.

V. Conclusions

The lack of an authoritative single source of xAPI recipe truth will demotivate the universal adoption of a consistent set of recipes. The lack of consistency in turn risks driving up the costs and decreasing the ability of collecting well-understood data across organization boundaries. The negative impact of this inconsistency will increase as learning analytics services aggregate to regional and national services. The xAPI recipes need to be seen in their infrastructural context. An ETL layer with communal best practices encoded in the transforms and applied across the higher education sector can enforce the authoritative standard and decrease overall costs. Further, synthetic data is useful in experimenting with services before the data governance is fully in place or as the legal and ethical culture adapts to technological realities.

We conclude with a rallying call to the learning analytics community: As part of the continued conversation surrounding the OLA initiative[16], the Apereo Foundation with support from SoLAR should consider acting as a safe location for synchronous curation of an authoritative set of xAPI recipes and ETL transforms if we wish to work with consistent data across a wide range of organizations and to later deploy consistent interventions for students triggered by predictive models using this data.

Acknowledgment

The efforts of M. Scheffel, H. Drachsler and S. Ternier have partly been funded by the ECO project (grant no. 21127) and the LACE project (grant no. 619424).

References

[1] A. Downes, A. Shahrazad, R. Smith, "Sharing between LRSs: a collaborative experiment in practical interoperability," March 2015, retrieved from http://experienceapi.com/wp-content/uploads/2015/03/whitepaper.pdf

[2] K. Kitto, S. Cross, Z. Waters, and M. Lupton, "Learning analytics beyond the LMS: the connected learning analytics toolkit," Proc. Fifth Int. Conf. on Learning Analytics and Knowledge (LAK'15), ACM, New York, NY, USA, March 2015, pp. 11–15, doi:10.1145/2723576.2723627.

[3] IMS Global Learning Consortium, "Learning measurement for analytics whitepaper," 2013, retrieved from https://www.imsglobal.org/sites/default/files/caliper/IMSLearningAnalyticsWP.pdf

[4] F. Brouns, J. Mota, L. Morgado, D. Jansen, S. Fano, A. Silva, and A. Teixeira, "A networked learning framework for effective MOOC design: the ECO project approach," in Proc. 8th EDEN Research Workshop, A. M. Teixeira, & A. Szücs, Eds. EDEN, Budapest, Hungary, 2014, pp. 161–171.

[5] D. Di Mitri, M. Scheffel, H. Drachsler, D. Börner, S. Ternier, and M. Specht, "Learning pulse: using wearable biosensors and learning analytics to investigate and predict learning success in self-regulated learning," in Proc. CrossLAK 2016 Workshop, to appear.

[6] C. Qu, and W. Nejdl, "Towards interoperability and reusability of learning resources: a SCORM-conformant courseware for computer science education,". Proc. 2nd IEEE Int. Conf. on Advanced Learning Technologies (ICALT'02), Kazan, Tatarstan, Russia, September 2002.

[7] H.-C. Schmitz, M. Wolpers, U. Kirschenmann, and K. Niemann, "Contextualized attention metadata," in Human Attention in Digital Environments, C. Roda, Ed. Cambridge University Press, Cambridge, UK, 2011.

[8] N. Manouselis, H. Drachsler, K. Verbert, and E. Duval, Eds. Recommender Systems for Learning, Springer, Berlin, 2012.

[9] S. Dietze, H. Drachsler, and D. Giordano, "A survey on linked data and the social web as facilitators for TEL recommender systems," in Recommender Systems for Technology Enhanced Learning: Research Trends & Applications, N. Manouselis, K. Verbert, H. Drachsler, and O.C. Santos, Eds. Springer, New York, 2014, pp. 47–77.

[10] P. Berking, "Choosing a learning record store (LRS)," November 2015, retrieved from http://adlnet.gov/adl-assets/uploads/2015/11/Choosing-an-LRS.pdf

[11] D. Tamraparni, and T. Johnson, Exploratory Data Mining and Data Cleaning, Wiley, 2003.

[12] A. Bakharia, K. Kitto, A. Pardo, D. Gašević, and S. Dawson, "Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit," in Proc. of the Sixth Int. Conf. on Learning Analytics and Knowledge (LAK'16). ACM, New York, NY, USA, April 2016, pp. 378-382, doi:10.1145/2883851.2883882.

[13] R. Koper, and B. Olivier, "Representing the learning design of units of learning," Educational Technology & Society, vol. 7(3), July 2004, pp. 97–111.

[14] A. Berg, M. Scheffel, H. Drachsler, S. Ternier, and M. Specht, "The Dutch xAPI experience," in Proc. of the Sixth Int. Conf. on Learning Analytics and Knowledge (LAK'16). ACM, New York, NY, USA, April 2016, pp. 544-545, doi:10.1145/2883851.2883968.

[15] https://solaresearch.org

[16] https://solaresearch.org/initiatives/ola/